50th Year
RAND
A NATIONAL RESOURCE
1948~1998

Muddy Learning: Evaluating Learning in Multi-User Computer-Based Environments

David McArthur, Matthew Lewis

DRU-1688-DARPA

February 1998

Prepared for the Defense Advanced Research Projects Agency

National Defense Research Institute

Preface

The Defense Advanced Research Projects Agency (DARPA) recently completed Computer-Aided Education and Training Initiative (CAETI), an ambitious program to develop and evaluate innovative educational technologies. This program not only supported the development of new educational software prototypes, but also implemented and field-tested them in several Department of Defense Dependent Schools (DoDDS). To support this effort, RAND undertook a small project to analyze some of the complexities associated with measuring the learning benefits of, and resolving the implementation challenges to, a particularly novel class of learning technologies. The results of this project have been briefed to the CAETI program manager; this short paper summarizes these conclusions.

This research was conducted for DARPA within the Acquisition and Technology Policy Center of RAND's National Defense Research Institute, a federally funded research and development center sponsored by the Office of the Secretary of Defense, the Joint Staff, the unified commands, and the defense agencies.

Table of Contents

Preface iii

Introduction 1

Why it is tough to evaluate MUDs for learning 1

Evaluating what students might learn with MUDs 3

Beyond "What is learned" 7

Introduction

Computer-based environments where many users interact in real-time are growing increasingly popular, especially as more people gain access to networks like the Internet. One class of such technologies, known variously as MUDs (Multiple User Dimension or Multi-User Dungeons and Dragons, take your pick), MUSEs (Multiple User Synthetic Environments), and MOOs (Multi-User Object Oriented), enables users to create new "rooms" in virtual worlds, define their own personnaes, and engage visitors in rich dialogues. Most MUDs[1], especially the earliest ones (some have been evolving for well over a decade), are text-based; however many now incorporate graphics, as network tools increase in sophistication and bandwidth to support the demands of multi-media. At the same time, MUDs have started to expand their market niche. Previously used mainly as "chat rooms" for social interaction or as programming environments for creating new rooms, many developers are now seriously considering how MUDs might provide novel *educational venues*. In this paper we consider briefly some claims about the possible educational benefits of MUDs, and the challenges of evaluating MUDs from an educational perspective.

Why it is tough to evaluate MUDs for learning

Evaluating the impact of a new learning technology is always challenging. The simplest and most familiar kind of evaluation often looks like a "horse race". One technology is pitted against another by arranging for two classrooms that are otherwise similar to use the different tools; the technology whose classroom does better – usually on some standardized test – wins. If the winning technology is a challenger to some existing method of teaching (say, for example, an intelligent tutoring system for algebra in contrast to traditional text-

[1]In this paper we will use the term "MUD" exclusively, with the understanding that it is intended generically to encompass MOOs, MUSEs, MUSHes and related worlds.

based methods) then the challenger can claim to be an improved way to help students learn.

"Horse race" evaluations are compelling, if you can conduct them properly. Unfortunately, for most MUDs, this will be impossible, simply because they do not meet the stringent constraints on a valid horse race evaluation. Such evaluations really make sense when the *only* thing one wants to do with a new learning technology is change *how* students learn, and hence when the central purpose of evaluation is to see if this new technology enhances students' learning outcomes, according to some accepted test or measure. If other things also change, in effect, the horses are not running on the same track, and so test results do not permit a simple comparison of the two learning technologies.

For the most part, developers of educational MUDs are not just trying to develop technologies that change how students learn; in fact several different kinds of changes are often associated with MUDs:

- *How students learn.* Not only are MUDs technology-intensive methods of learning; but, because they are rooted in networked communication, they also emphasize a collaborative and cooperative perspective on knowledge acquisition.
- *What students learn.* While some MUDs claim to help students learn traditional school subjects (e.g., reading) many focus on personal and social learning outcomes, in addition to (or sometimes in contrast to) academic learning.
- *Where students learn.* A few MUDs have recently moved into classrooms, but they remain much more popular in informal, learning venues – in labs, at home, or wherever students can access a network.
- *What learning and evaluation are about.* Perhaps the deepest difference between MUDs and traditional tools for learning – including most computer-based systems – is philosophical. Developers often root their

MUD designs in a constructive epistemology and theory of learning; they also frequently advocate a situated approach to evaluation, believing that traditional "atomistic" evaluation strategies are not only obsolete but fundamentally flawed.

Of course, with some or all of these changes happening simultaneously, a MUD can hardly be assessed using simple instruments like horse-race evaluations. But the challenges to evaluating MUDs go deeper than just replacing our familiar assessment tools with newer and better ones. Today, few developers have begun to articulate the kinds of educational changes they want their MUD to support – the changes we just listed come from our (relatively casual) analysis of some existing MUDs, not from the literature. Similarly, *educational goals* are often only tacitly associated with MUDs, not explicitly announced; and *evaluation purposes* and *questions* also usually remain implicit. All these issues must be untangled before we can begin to craft specific instruments appropriate to evaluating MUDs for learning and education.

Evaluating what students might learn with MUDs: Some initial thoughts

In this brief paper we can only begin to address a few of these challenges. We will look, in particular, at some claims concerning what students might learn with MUDs, then discuss evaluation questions that follow from these claims, and, finally, we review the types of assessment tools that could answer these questions. Our discussion is organized around Table 1. This table was not constructed by reviewing the MUDs literature, but rather through informal discussions with MUD-developers at MudShop II.[2] So, in a sense, it represents a

[2] "MUDshop II" was a workshop on collaborative learning environments, hosted by Dr. Kirstie Bellman and DARPA, during September 1995 in San Diego CA. Matthew Lewis from RAND participated in the workshop and lead the group on assessment issues. Many insights in this paper stem from his participation.

relatively direct attempt (from the horses' mouths, so to speak) to articulate claims concerning the different learning outcomes MUDs might foster.

Knowledge type	*What is learned?*	*Examples*	*Assessment Tools*
Academic skills (traditional)	• deeper understanding • programming • qualitative modeling • quantitative modeling • domain-specific knowledge • reading/comprehension • writing/communication • scientific method • teaching/mentoring	• learning history (*Rebel! MUD*) •(*LambdaMOO*) • collectively construct a rainforest •math MUD with stored word-problems • business, Egyptology, math • multi-user electronic communication (*several MUDs)* • collaborative data collection and analysis • using network helps learners become teachers	• transfer tasks; teach topic • traditional programming tests; on-line versions • delayed access; reconstruction • traditional transfer; on-line tests • traditional tests; on-line versions • traditional test; on-line versions • traditional; hands-on • observation; peer assessment • observation; peer assessment
Meta-cognitive skills	• learning to learn: -problem finding -information filtering -self-diagnosis -help-seeking • integrating information • creativity	-finding useful questions -separating wheat from chaff -evaluating own knowledge -know how to get support	-protocol analysis -protocol analysis -protocol analysis -protocol analysis • synthetic questions for integration
Social skills	• communication skills • interpersonal interaction • collaboration • leadership	• writing and speaking skills • "How not to be a jerk" • working well with others • able to provide guidance	• peer, teacher reports; "tattling" • observation; peer report • analysis of comm. patterns • observation
Personal skills	• self-esteem • willingness to interact • empowerment • tolerance • trust of others • enjoyment of learning • personal discipline	• belief in self-worth • overcome shyness (*DragonMud*) • belief in personal impact • accept behavior of others • willing to rely on others • desire to acquire skills • set and accomplish goals	• existing commercial • observation • self-report; others' reports • self-report; observation • self-report; observation • self-report; observation • observation
Other general benefits	• increased engagement/ motivation • raised expectations • community-building • tailorable to different types of learners	• "into learning" • aspire to higher education/training • care for environment and others	• time on task; attendance • graduation rate; post-grad path • self-report; others' reports

Table 1 – Different claims for what can be learned in MUDs.

The first two columns of Table 1 outline an extensive list of the kinds of knowledge students might acquire in MUDs; the third column offers some examples of skills, or refers to specific MUDs that might facilitate such learning (in the rather few cases where existing systems are already showing educational effects); and the last column lists ways this learning might be assessed.

Table 1 is clearly very rough, reflecting its origins in informal brainstorming sessions. The five kinds of learning outcomes, for example, probably should be refined, and you also can question whether a specific skill belongs in one category or another. (Why is increased motivation a general benefit rather than a personal skill?). However the skills are sorted, though, the main point is clear: advocates are claiming that well-crafted MUDs can help learners acquire many kinds of skills and very diverse ones. Some, especially in the academic category, correspond quite closely to those taught in traditional classrooms; but most skills in the other categories are not allied with standard school subjects. Similarly, some skills are relatively well-defined (mainly those also taught in schools) while others are rather ill-defined (mainly those not taught in schools). At the very least, the diversity of suggested learning outcomes means that new instruments will be needed to measure them; the novelty of some claims means that creating such assessment tools may be very challenging. We discuss a few of these instruments in the following paragraphs – keeping in mind that the list of assessment tools in Table 1 was generated in the same informal MudShop II conversations that lead to the list of learning outcomes.

Academic skills. These are closest to traditional school outcomes, and so it is hardly surprising that the assessment tools to measure them, borrowing from classroom experience, are familiar and well-defined (see the assessment tools column in the academic skills row of Table 1). For example, the same standardized tests we use in algebra classrooms can measure learning in a math MUD. Even here, though, evaluation can look very different than classroom

assessment. Math MUD tests can at least be put on-line. More interestingly, tests may be integrated with the ongoing tasks and challenges of moving through different MUD rooms (for example, imagine hunts where access to a room was contingent on answering a math question).

Meta-cognitive skills. The open-ended structure of MUDs offers students ample opportunities to learn problem-finding and other reflective skills. However, these meta-cognitive abilities also are increasingly the focus of broad educational reform efforts; some of the instruments now under development by reform groups can be used to assess MUDs as well. As we note in the Table (under assessment tools for meta-cognitive skills), various styles of protocol analysis are often used to uncover meta-cognitive skills. Here too, however, traditional instruments might be creatively integrated into the structure of a MUD, rather than used as a adjunct assessment tool. For example, the raw material for protocol analysis – the transcript of student behaviors – can be collected automatically, as an audit trail of students' interactions with MUD objects and other MUD participants. It should be possible to develop "agents" that at least partially automate the analysis of these protocols.

Social and personal skills. Most MUD developers believe the personal and social learning MUDs can foster are among their most important benefits. Some even claim that persistent, network-based environments, where users define their own personnae as well as construct virtual shared spaces, can capture all the key features of enculturation. Even disregarding the most controversial (and most interesting) claims, almost everyone would agree skills like tolerance and collaboration are among the most important ones to acquire. But they are not ones taught explicitly in school, and, consequently, the assessment tools we listed to measure them (mainly observation and self-reports) are relatively vague and unreliable. They need to be improved. As with academic and meta-cognitive skills, some innovative assessment tools may be integrated into the structure of the MUD itself; for example growth of leadership skills within a

MUD could be measured by comparing the quantity and quality of questions students *ask* of other users to ones they *answer*.

Beyond "What is learned": additional challenges to evaluating MUDs

The previous section suggested that many learning goals associated with MUDs remain unclear, that evaluation questions addressing those goals need to be sharpened, and that tools to answer the questions require much further thought. But now let's step into the future and assume we have been able to refine all of the learning outcomes, and to implement all of the assessment tools we need. Would evaluating a MUD then just amount to listing its learning goals, grabbing the tools associated with these goals, and turning them loose in the classroom? Not necessarily; and reviewing a few of the reasons why will give us a deeper understanding both of evaluation and of MUDs.

The first reason evaluation will not be this simple is hinted at by the last general benefit mentioned in Table 1: that MUDs could be tailored to different populations of users. This flexibility is clearly desirable; but at the very least it means that a MUD can be associated with different learning goals – not a single fixed set – depending on the needs of students. More broadly, although some MUDs are purpose-built for specific curricula (e.g., *Rebel!* helps students learn American history), most are designed as *open systems.* As such, they are intended, within limits, to be appropriated by each classroom (or other learning venue) for its own educational purposes. In part, this is why developers of MUDs talk of nurturing the *evolution* of on-line communities and cultures.

But if the purposes and educational goals of a MUD evolve over time, then it is naive to assume we can determine the evaluation questions for the MUD at the outset, let alone the appropriate assessment tools. These too may need to evolve – through repeated discussions with the MUD developers, classroom members, and evaluation specialists – as the MUD takes shape. Perhaps effective evaluation strategies will simply pick and choose from among our

existing repertoire of assessment tools, depending on how the MUD's purposes evolve. But new tools may also need to be constructed on-demand, to address the MUD's changing learning goals, and to adapt to its structure. In any event, the flexibility of MUDs adds greatly to the challenges of evaluating them.